International Journal of Technology (2015) 6: 961-970
ISSN 2086-9614

BAYESIAN MIXTURE MODEL FOR PREDICTION OF BUS ARRIVAL TIME

Misbahuddin[1,2*], Riri Fitri Sari[1]

[1] *Department of Electrical Engineering, Faculty of Engineering, University of Indonesia, Kampus UI Depok, Depok 16424, Indonesia*
[2] *Department of Electrical Engineering, Faculty of Engineering University of Mataram, Mataram 83121, Indonesia*

(Received: June 2015 / Revised: September 2015 / Accepted: September 2015)

ABSTRACT

Providing travelers with accurate bus arrival time is an essential need to plan their traveling and reduce long waiting time for buses. In this paper, we proposed a new approach based on a Bayesian mixture model for the prediction. The Gaussian mixture model (GMM) was used as the joint probability density function of the Bayesian network to formulate the conditional probability. Furthermore, the Expectation maximization (EM) Algorithm was also used to estimate the new parameters of the GMM through an iterative method to obtain the maximum likelihood estimation (MLE) as a convergence of the algorithm. The performance of the prediction model was tested in the bus lanes in the University of Indonesia. The results show that the model can be a potential model to predict effectively the bus arrival time.

Keywords: Arrival time prediction; Bayesian network; Gaussian mixture model

1. INTRODUCTION

In developing countries, the problem of traffic congestion becomes daily routine due to the increase of traffic volume characterized by slower speed, increased vehicular queuing, and longer trip times. The impact of the traffic congestion is the lost time and the wasted fuel. Furthermore, emission produced by pollution is clearly dangerous for human and environment. One way for overcoming the problem is improving the infrastructure capacities to cope with the increasing number of vehicles. However, this solution has a limitation, so that other alternative options are necessary to be explored to anticipate the growing traffic demand. The answer to such a problem is the utilization of the Intelligent Transportation System (ITS).

There are two types of technology in the intelligent transportation system to handle the transportation management, i.e. Advance Traveler Information Systems (ATIS) and Advance Traffic Management Systems (ATMS). These technologies have become hot topics in the last decade. Many researchers have devoted their time in this research (Chen et al., 2012; Haitao et al., 2013; Jeong & Rilett, 2004; Pengfei et al., 2014; Yu et al., 2011). They have shown that providing the information of the bus arrival time is an influential element to improve the performance of the ATIS and the ATMS. Therefore, the prediction of the bus arrival time is an important issue in reducing the traffic congestion and the CO_2 emissions.

One of the most important applications of the ITS is the Advanced Public Transportation System (APTS) that has provided the real time traveler information. The APTS can contribute in reducing congestion by diverting more travellers to use public transportation. Rather than,

* Corresponding author's email: misbah@te.ftunram.ac.id, Tel. +62-370-636126, Fax. +62-370-636523
Permalink/DOI: http://dx.doi.org/10.14716/ijtech.v6i6.1350

one of the most important information provided by the APTS is a prediction of accurate bus arrival time for travelers to reduce the long waiting time.

In the last decade, there are various methods in the intelligent transportation system. The methods can be divided into four categories: (1) statistic methods (regression analysis and historical average models); (2) Kalman filter models; (3) machine learning approaches (artificial neural network, k-nearest neighbor, and support vector machines); and (4) other methods (Markov chains, and passengers' participatory sensing).

(Tongyu et al., 2012) used two data types, i.e. historical global position system (GPS), and the automatic fare collection impact of signalized intersection (AFC) data, to establish two artificial neural network (ANN) model to predict the real time bus arrival. Their results showed that the ANN model can outperform the other two models in most scenarios. (Tao et al., 2012) applied the modified k-nearest neighbor (k-NN) method integrated into the cluster analysis and principal component analysis for predicting the arrival time, using historical GPS data. The results showed that the k-NN method has the accuracy of the prediction lower than that of the ANN model. (Yu et al., 2011) found that the support vector machine (SVM) is the best model for predicting the bus arrival time based on the running times of buses on multi routes, which are more accurate than that of single-route models. Moreover, Pengfei et al. (2014) presented a model prediction for the bus arrival time based on the bus passengers' participatory sensing. Their evaluation result shows that the performance in the system is more accurate than that of the system with buses equipped with the GPS.

In a congestion network, the prediction of the bus travel time has a high uncertainty. The robust and effective model for an uncertainty problem is the ANN model. The previous study results show that the ANN is one of the most efficient models for predicting the bus arrival time (Tongyu et al., 2012). Therefore, the weak point of the ANN is the long training time (Lingli et al., 2013). Generally, the weakness of the available methods is when an incomplete data that is caused by the malfunctions of the recording system or the error measurement in the data collection is used for predicting (Shiliang et al., 2006).

This paper proposes a model through the development of Bayesian network with a Gaussian mixture model as the joint probability density function. Based on the fundamental theory of the Bayesian network, which studied comprehensively by the researchers of statistical analysis, artificial intelligence, and machine learning, has been proven as a robust model to address the uncertainty problems such as in predict or forecast (Shiliang et al., 2006). In this paper, we focus on conducting traffic state modeling and prediction of the bus arrival time. We review the work conducted by Lingli et al., 2013 and Shiliang et al., 2006. Both works are different from our model. Shiliang et al. (2006) used Bayesian network model to forecast the traffic flow in a line. Meanwhile, Lingli et al. (2013) just utilized Bayesian network to predict the bus travel time with the state transition matrix generation as a joint probability of the Bayesian network.

Our model provides two main contributions that consist of (i) the usage of the Bayesian network with Gaussian mixture model as a joint probability density to model a new traffic state considering the links of various directions; (ii) the usage of the Nokia X as the Global Positioning System (GPS) device to gather the row data from each bus to be stored in the database of the backend server. Moreover, the Nokia X smartphone is also used as a transmitter to send the position along with timestamp of the bus.

The remainder of this paper is organized as follows: we continue to introduce the Bayesian networks and parameter estimation of the Gaussian mixture model as a joint probability density in Section II. Section III describes the model construction and experiment to obtain the final results. Finally, Section IV concludes the paper and the discusses some direction of the future work.

2. METHODOLOGY

2.1. Bayesian Network

A Bayesian network, also known as probabilistic network, is formally a pair B=(G, Θ) (Pernkopf et al., 2012) where G is a directed acyclic graph, and Θ is a set of parameters in all conditional probability distributions. The graph G consists of nodes and edges, where nodes represent random variable, and edges represent directed dependencies. The conditional distribution determines the strength of the association between the parent and child nodes. In the Bayesian network, a joint probability distribution is a set of variable $x_{1:N}$. it is a product that is formed by the chain rule (Shiliang et al., 2006):

$$P(x_{1:N}) = \prod_{i=1}^{N} P(x_i | p_{x_i}) \tag{1}$$

where p_{x_i} are the set parents of the variable x_i according to a given Bayesian network (p_{X_i} can be empty if node *i* has no parent).

2.1.1. Parameter estimation of Gaussian Mixture

Inference Bayesian network is a reasoning process that begins from the causal nodes to a target node for predicting an expected value in the Bayesian network's output. In a Bayesian network, the joint probability density represents the relationship between the input and the output. In our research, we adopt the Gaussian Mixture Model (GMM) as the joint probability density function of the Bayesian network. To formulate, we assume that a finite data set X = { x^1, ..., x^N} is a multidimentional random vector. The GMM gives the probability density in a mixture of K Gaussians as follows.

$$P(x \mid \theta) = \sum_{j=1}^{K} w_j \, N(x \mid \theta_j) \tag{2}$$

where w_j is the prior probability (weight) of the j^{th} Gaussian such that $\sum_{j=1}^{K} w_j = 1$ and the parameters of the Gaussian probability density are represented by $\theta_j = (\mu_j, \Sigma_j)$.

To estimate the parameters θ_j of the GMM that fits the data, we use the Expectation Maximization (EM) (Roberts et al., 1998) algorithm. The algorithm, one of the most popular approaches to maximize the likehood, iterates in two steps. The first step, E-Step estimates the distribution of the hidden variable given the data and the current value of the parameters. In second step, M-Step maximizes the joint distribution of data and the hidden variable. On each Expectation Maximum (EM) iteration, the following iterative equations that guarantee a monotonic increase in the model's likehood value are given as (Roberts et al., 1998).

$$w_j = \frac{1}{N} \sum_{i=1}^{N} p(j \mid x_{ki}, \Theta^t)$$

$$\mu_j = \frac{\sum_{i=1}^{N} p(j \mid x_i, \Theta^t) x_i}{\sum_{i=1}^{N} p(j \mid x_i, \Theta^t)}$$

$$\Sigma_j = \frac{\sum_{i=1}^{N} p(j|x_i, \Theta^t)\,(x_i - \mu_j)(x_i - \mu_j)^T}{\sum_{i=1}^{N} p(j \mid x_i, \Theta^t)} \tag{3}$$

where, N is the size of data set X and j = 1, ..., K.

2.1.2. Expected value in target node of Bayesian Network

According to Bayesian's rule, expected value of the node target can be obtained by inferecing in the Bayesian network, which is derived through the conditional probability density function. The Equation 4 denotes the conditional probability with the GMM as joint probability density (Shiliang et al., 2006).

$$\begin{aligned} P(x_H|x_E) &= \frac{P(x_E|x_H)\,P(x_H)}{P(x_E)} \\ &= \frac{\sum_{j=1}^{K} w_j\, N\left(x_E;\mu_{jE},\Sigma_{jEE}\right) N(x_H;\mu_{jH|E},\Sigma_{jH|E})}{\sum_{i=1}^{K} N(x_E;\mu_{jE},\Sigma_{jEE})} \\ &= \sum_{j=1}^{K} \delta_j\, N(x_H;\mu_{jH|E},\Sigma_{jH|E}) \end{aligned} \tag{4}$$

where

$$\begin{aligned} \delta_j &= \frac{w_j\, N\left(x_E;\mu_{jE},\Sigma_{jEE}\right)}{\sum_{i=1}^{K} N(x_E;\mu_{jE},\Sigma_{jEE})} \\ \mu_{jH|E} &= \mu_{jH} - \Sigma_{jHE}\Sigma_{jHH}^{-1}(\mu_{jE} - x_E) \\ \Sigma_{jH|E} &= \Sigma_{jHH} - \Sigma_{jHE}\Sigma_{jEE}^{-1}\Sigma_{jEH} \end{aligned} \tag{5}$$

The predicted value can be obtained from the target node of the Bayesian network in Equation 6 as follows:

$$\begin{aligned} x_H &= E\,(x_H|x_E) \\ &= \int P(x_H|x_E)x_H dx_H \\ &= \sum_{j=1}^{K} \delta_j \int N\left(x_H;\ \mu_{jH|E},\Sigma_{jH|E}\right)x_H dx_H \\ &= \sum_{j=1}^{K} \delta_j\, \mu_{jH|E} \end{aligned} \tag{6}$$

2.2. Model Construction and Experiments

2.2.1. Construction of Bayesian Network

The prediction for the bus arrival time is a problem related to time series of traffic state (vehicle flows) in each road link. The Bayesian network approach should consider the time factor of the traffic state as well. To design the Bayesian network, we need to construct the relationship between the cause nodes (parent nodes) and a target node. In this study, before predicting the bus arrival time for the next bus stop, we need to get the traffic state in the upstream links, downstream links, and target links respectively at the latest and current time. The traffic state is the vehicle's average speed that is updated each 5 minutes.

To construct the structure of the Bayesian network, we should refer to the road network. Figure 1 shows an example of a road network. The circles A and B mean a crossroad and the arrows that symbolized A1, A2, and A3 represent the direction of the upstream link. Furthermore, the symbol of B1, B2, and B3 are the directions of the downstream links.

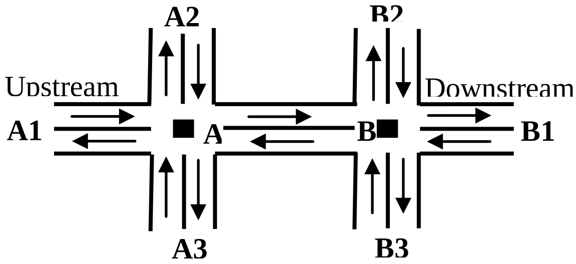

Figure 1 A road network example

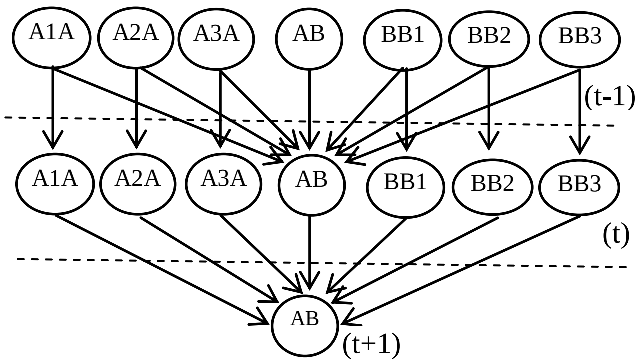

Figure 2 The Bayesian network between the observed link AB and its neighbour's links

Figure 2 shows a structure of the Bayesian network according to the example of road network. The target node AB_{t+1} represents the predicted traffic state on the AB link at after time t. It depends on the cause (parent) nodes, $A1A_{(t)}$, $A2A_{(t)}$, $A3A_{(t)}$, $BB1_{(t)}$, $BB2_{(t)}$, $BB3_{(t)}$, and $BC_{(t)}$ at the current time. The $BC_{(t)}$ node, which is also one of the parent nodes of the target node, depends on $A1A_{(t-1)}$, $A2A_{(t-1)}$, $A3A_{(t-1)}$, $BB1_{(t-1)}$, $BB2_{(t-1)}$, $BB3_{(t-1)}$, and $BC_{(t-1)}$ nodes at before time t.

The joint probability of the Bayesian network in Figure 2 can be written as follows.

$$
\begin{aligned}
&P((A_iA)_{(t-1)}, AB_{(t-1)}, (BB_i)_{(t-1)}, (A_iA)_{(t)}, AB_{(t)}, (BB_i)_{(t)}, AB_{(t+1)}) \\
&= \prod_{i=1}^{3} P\big((A_iA)_{(t)} \big| (A_iA)_{(t-1)}\big) \times \prod_{i=1}^{3} P\big((BB_i)_{(t)} \big| (BB_i)_{(t-1)}\big) \\
&\quad \times \prod_{i=1}^{3} P\big(AB_{(t)} \big| (A_iA)_{(t-1)}, AB_{(t-1)}, (BB_i)_{(t-1)}\big) \\
&\quad \times \prod_{i=1}^{3} P\big(AB_{(t+1)} \big| (A_iA)_{(t)}, AB_{(t)}, (BB_i)_{(t)}\big)
\end{aligned} \tag{7}
$$

2.2.2. *Prediction framework*

The output of the Bayesian network in the Equation 6 is a prediction result of the traffic state in an observed link at a time t+1. In this case, we know that the traffic state represents the vehicle's average speed. For example, the link AB in Figure 1 as an observed link, and therefore the travel time can be obtained as follows.

$$t_{r(AB)} = \frac{D_{AB}}{\bar{v}_{AB}} \tag{8}$$

where $t_{r(AB)}$ is travel time from location A to B, D_{AB} is the distance between point A and B, and $\bar{v}_{Ab}$ is the average speed in the link AB.

If we need to predict the bus travel time from a certain location toward an end bus stop, the bus will pass some bus stops as illustrated in Figure 3. The equation was not as simple as the Equation 8. Assume that bus 101 is located at the location A1 toward the target bus stop at the location B1, which passes some bus stops, the travel time can be expressed as:

$$t_{r(A1B1)} = \frac{D_{A1A}}{\bar{v}_{A1A}} + \sum_{i=1}^{n} \frac{D_i}{\bar{v}_i} \tag{9}$$

where D_{A1A} is the distance between the location A1 and the first bus stop in the link A1A, and D_i is the distance between the bus stops that have been passed by the bus until the nth target bus stop on the location B1.

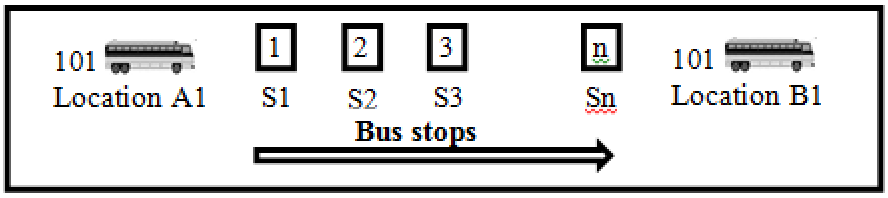

Figure 3 The illustration of the bus 101 that passes through some bus stops until the nth target bus stop (location B1)

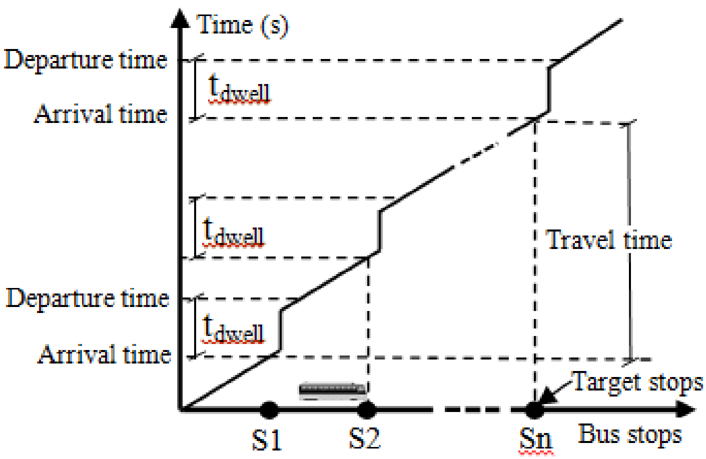

Figure 4 The dwell time at a bus stop (Tongyu et al., 2012)

The dwell time of any bus stop i that starts on the arrive of the bus and ends on the departure of the bus at the bus stop. The dwell time t_{dwell} is illustrated in Figure 4 and can be calculated as follows (Jian et al., 2013):

$$t_{dwell} = \sum_{i=1}^{N-1} \frac{t_d - t_a}{N} \tag{10}$$

Finally, the bus arrival time can be obtained as follows.

$$t_{BAT} = T_{A1} + (t_r + t_{dwell}) \tag{11}$$

where T_{A1} is the arrival time of the bus at the location A1.

3. RESULTS AND DISCUSSION

In this research, the bus lanes in the University of Indonesia are used as a test bed for the algorithm application object. The bus lanes contain of two routes, red route and blue route that opposite directions. The bus lanes can be seen in Figure 5 where the star marks are the bus stops that contain of 16 bus stops. The origin and destination of the bus are the bus stop at the Wisma Makara on the top of the map. The buses run around the cycle's route of the campus of the University of Indonesia in two opposite directions.

The Automatic Vehicle Location (AVL) system shown in Figure 6 is used to gather the row data from each bus to be stored in the database of the backend server. Each bus is equipped with a Nokia X smartphone that acts as a Global Positioning System (GPS) device. Moreover, the Nokia X also sends the bus's data (id bus, time stamp, and position) every five seconds to the backend server. The row data is captured from the October 1 to 31, 2014, totaling 23 days (five business days in a week). The total number of the row data are 3450 units. One unit of the row data consists 15 sample points. Thus, there are 3450 units × 15 sample points = 51750 sample points. In our experiment, we divided the sample points into two parts. One part is used as the training set and the other is used as a testing set. The training set is used to estimate the parameters of the GMM.

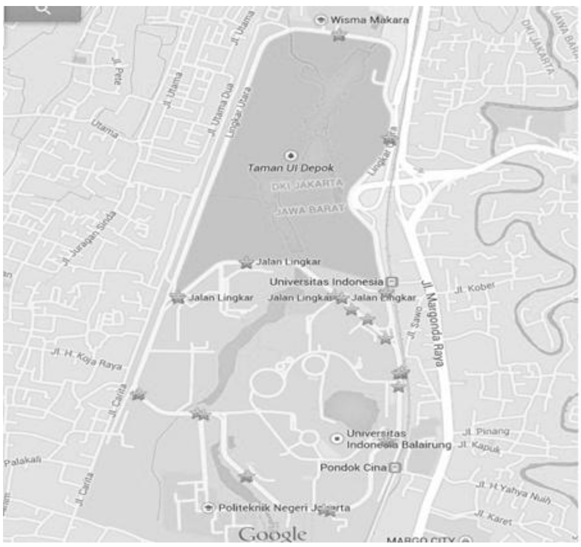

Figure 5. The bus lanes in the University of Indonesia

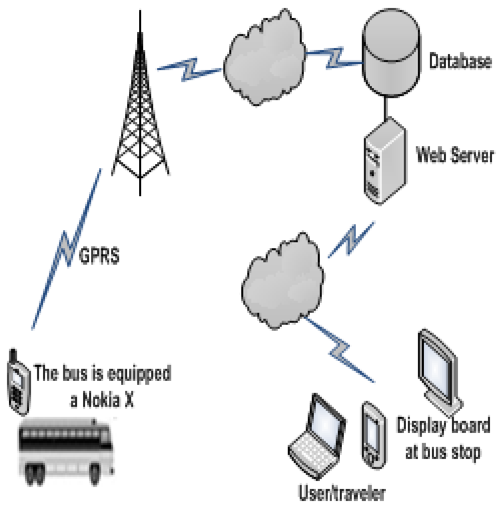

Figure 6. The Automatic Vehicle Location (AVL)

To evaluate the accuracy of our model, Figure 7 shows a comparison between the predicted values and the real values of travel time for the bus 101 in the red route. The testing was conducted 10 times for a certain link.

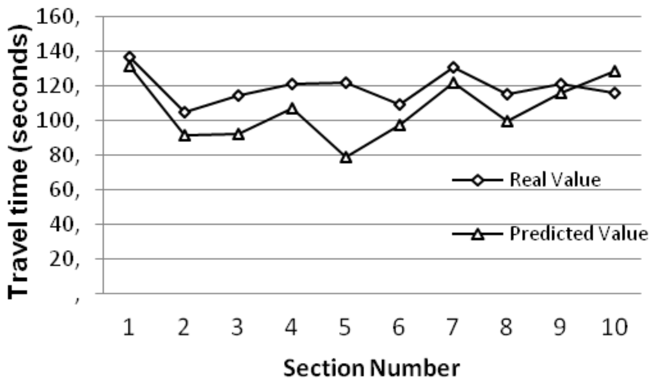

Figure 7 The real values and the predicted values in the comparison of bus travel time

The predicted performance of the our approach is evaluated by four kinds of error indices. These indices are expressed respectively in Equations 10, 11, 12, and 13 as follows:

Mean Absolute Percentage Error :
$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|RV - PV|}{RV} \times 100\% \tag{10}$$

Mean Absolute Error :
$$MAE = \sum_{i=1}^{N}\frac{|RV - PV|}{N} \tag{11}$$

Maximum Relative Error :
$$MXRE = \max\left(\frac{|RV - PV|}{RV} \times 100\%\right) \tag{12}$$

Maximum Absolute Error :
$$MXAE = \max|RV - PV| \tag{13}$$

where RV is the real value of the bus travel time and PV is the predicted value of the bus travel time.

After calculating the travel time as shown in Figure 7, we can assess the overall predicting performance of our model using four error indices. Table 1 shows the error value for the four error indices. Furthermore, Figures 8 and 9 show the graph of MAPE and MAE respectively. Both are also important information to assess the best high error point. In this case, we use two error indices, MXRE and MXAE, to show the best high error point.

Table 1 Error value of the four indices

Error Indices	Error value
MAPE (%)	12.80
MAE	15.06
MXRE (%)	35.47
MXAE	43.20

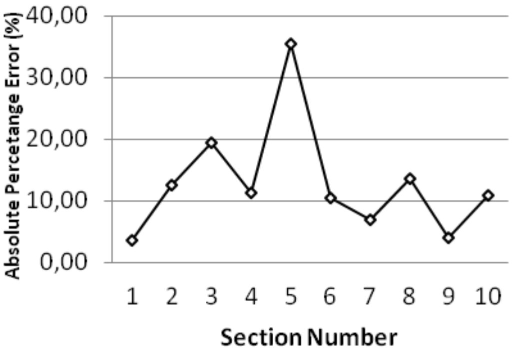

Figure 8 The absolute percentage error in each section

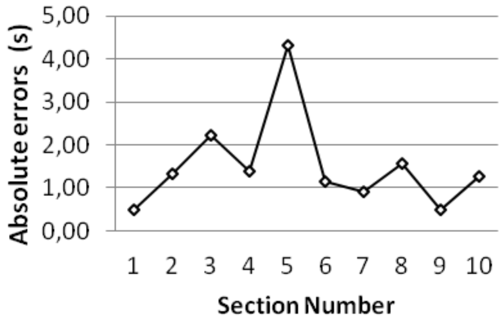

Figure 9 The absolute errors in each section

Table 1 shows that both the MAPE (12.8%) and MAE (15.06) are two acceptable error indices. Although, two other error indices, which are the MXRE (35.47%) and MXAE (43.20) that occurs in section number 5 as shown in Figures 8 and 9 respectively, have a relatively high error value. These error values are caused by the error rate of the fairly high error values in MAPE and MAE. The following is the reasons of the error values:

1) The limited numbers of the raw data as the training data for the Bayesian network cause the system to be unable to accommodate the incidental occurrences in the road. With this regard, in the future work we will to enrich the raw data with other occurrences such as accidents, and congestions.
2) There are some blank spot points of GPRS signal on the bus routes of the University of Indonesia, so that the Nokia X smartphone sends incorrect data to the backend server.
3) The variable used as the influence factor to predict the bus arrival time only consider the average road speed of various directions as the observed road traffic state.

4. CONCLUSION

In this paper, we proposed a predicting model for the bus arrival time using Bayesian mixture model. One of the difficulties of the Bayesian network that uses the continuous variables is a proper way to obtain a joint probability distribution. For this intenstion, we use Gaussian mixture model (GMM) as a joint probability density of the Bayesian network. We use the Expectation Maximization (EM) Algorithm to estimate the new parameters of the GMM through an iterative method to obtain the maximum likelihood estimation (MLE) as a convergence of the algorithm. For experiment, we used the bus lanes in the University of Indonesia. To evaluate the accuracy of the model, we assess the performance with four kinds of the error indices. As the results, the MAPE is 12.8%, the MAE is 15.06, the MXRE is 35.47% for section number 5, and the MXAE is 43.20 for section number 5. These error indices are acceptable. In summary, the results shows that the model can be a potential model to predict the bus arrival time effectively.

5. REFERENCES

Chen, G., Yang, X., An, J., Zhang, D., 2012. Bus-arrival-time Prediction Models: Link-based and Section-based. *Journal of Transportation Engineering*, Volume 138(1), pp. 60–66

Haitao, Yu, Randong, Xiao, Yong, Du, Zhiying, He, 2013. A Bus-arrival Time Prediction Model Based on Historical Traffic Patterns. In: C*omputer Sciences and Applications (CSA), 2013 International Conference on Wuhan*

Jeong, R., Rilett, L.R., 2004. Bus Arrival Time Prediction using Artificial Neural Network Model. In: *Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference on Washington, D.C.*

Jian, D., Lu, Z., Yan, Z., 2013. Mixed Model for Prediction of Bus Arrival Times. In: *Evolutionary Computation (CEC), 2013 IEEE Congress on Cacun*

Lingli, D., Zhaocheng, H., Renxin, Z., 2013. The Bus Travel Time Prediction based on Bayesian Networks. In: *Information Technology and Applications (ITA), 2013 International Conference on Chengdu*

Pengfei, Z., Yuanqing, Z., Mo, L., 2014. How Long to Wait? Predicting Bus Arrival Time with Mobile Phone Based Participatory Sensing. *Mobile Computing, IEEE Transactions on*, Volume 13(6), pp. 1228–1241

Pernkopf, F., Wohlmayr, M., Tschiatschek, S., 2012. Maximum Margin Bayesian Network Classifiers. *IEEE Transactions Pattern Analysis and Machine Intelligence*, Volume 34(3), pp. 521–532

Roberts, S.J., Husmeier, D., Rezek, I., Penny, W., 1998. Bayesian Approaches to Gaussian Mixture Modeling. *IEEE Transactions Pattern Analysis and Machine Intelligence*, Volume 20(11), pp. 1133–1142

Shiliang, S., Changshui, Z., Guoqiang, Y., 2006. A Bayesian Network Approach to Traffic Flow Forecasting. *IEEE Transactions Intelligent Transportation Systems*, Volume 7(1), pp. 124–132

Tao, L., Jihui, M., Wei, G., Yue, S., Hu, N., 2012. Bus Arrival Time Prediction based on the k-Nearest Neighbor Method. In: *Computational Sciences and Optimization (CSO), 2012 Fifth International Joint Conference, Harbin, China*

Tongyu, Z., Jian, D., Jian, H., Songsong, P., Bowen, D., 2012. The Bus Arrival Time Service based on Dynamic Traffic Information. In: *The Application of Information and Communication Technologies (AICT), 2012 6th International Conference on Tbilisi*

Yu, B., Lam, W.H.K., Tam, M.L., 2011. Bus Arrival Time Prediction at Bus Stop with Multiple Routes. *Transportation Research Part C: Emerging Technologies*, Volume 19(6), pp. 1157–1170